A PATTERN SELECTION ALGORITHM IN KERNEL PCA APPLICATIONS

Ruixin Yang, John Tan, Menas Kafatos
Center for Earth Observing and Space Research (CEOSR)
College of Science
George Mason University
Fairfax, VA 22030, U.S.A.

Keywords: Data mining, knowledge acquisition, large-scale, dimension reduction.

Abstract: Principal Component Analysis (PCA) has been extensively used in different fields including earth science for spatial pattern identification. However, the intrinsic linear feature associated with standard PCA prevents scientists from detecting nonlinear structures. Kernel-based principal component analysis (KPCA), a recently emerging technique, provides a new approach for exploring and identifying nonlinear patterns in scientific data. In this paper, we recast KPCA in the commonly used PCA notation for earth science communities and demonstrate how to apply the KPCA technique into the analysis of earth science data sets. In such applications, a large number of principal components should be retained for studying the spatial patterns, while the variance cannot be quantitatively transferred from the feature space back into the input space. Therefore, we propose a KPCA pattern selection algorithm based on correlations with a given geophysical phenomenon. We demonstrate the algorithm with two widely used data sets in geophysical communities, namely the Normalized Difference Vegetation Index (NDVI) and the Southern Oscillation Index (SOI). The results indicate the new KPCA algorithm can reveal more significant details in spatial patterns than standard PCA.

1 INTRODUCTION

Principal Component Analysis (PCA) has been extensively used in different fields since introduced by Pearson in 1902 (as cited in (Von Storch and Zwiers, 1999)). This data decomposition procedure is known by various names in different disciplines such as Karhunen-Loève Transformation (KLT) in digital signal processing (Haddad and Parsons, 1991), Proper Orthogonal Decomposition (POD) in studies of turbulence coherent structure with nonlinear dynamical systems (Holmes et al., 1996), and Empirical Orthogonal Function (EOF) for one variable data (Lorenz, 1959) or Singular Value Decomposition (SVD) for multiple variables (Wallace et al., 1992) applied to earth science, in particular for climate studies.

In principle, PCA is a linear procedure to transform data for various purposes including dimension reduction (factor analysis), separation of variables, coherent structure identification, data compression (approximation), feature extraction, etc. Consequently, PCA results can be viewed and explained from various perspectives. One way to interpret the PCA results is to consider the PCA procedure projecting the original high dimensional data into a new coordinate system. In the new system, the space spanned by the first few principal axes captures most of the information (variances) of the original data (Krzanowski, 1988). Another point of view, commonly used in earth science, is to consider the PCA results of spatio-temporal data as a decomposition between the spatial components and temporal components. Once again, the approximation defined by the first few principal components gives the smallest total mean-square error compared to any other expansions with the same number of items (Emery and Thomson, 2001).

In earth science applications, the spatial components from the PCA decomposition are recognized as representative patterns because the spatial components are orthogonal to each other. Correspondingly, the uncorrelated time series (temporal components) are often used to study the relationships between the corresponding spatial patterns and a predetermined phenomenon such as the well-known El Niño, characterized by abnormally warm sea surface temperature (SST) over the eastern Pacific Ocean. Through this procedure, patterns can be associated with natural phenomena. One example of such association is found between Normalized Difference Vegetation Index (NDVI) patterns and ENSO (El Niño

Yang R., Tan J. and Kafatos M. (2006).
A PATTERN SELECTION ALGORITHM IN KERNEL PCA APPLICATIONS.
In *Proceedings of the First International Conference on Software and Data Technologies*, pages 195-202
DOI: 10.5220/0001320801950202

Southern Oscillation) by directly using spatial components of PCA (Li and Kafatos, 2000). Another well-known spatial pattern is obtained by regressing the leading principal time series from the sea-level-pressure (SLP) to the surface air temperature (SAT) field (Thompson and Wallace, 2000).

Although PCA is broadly used in many disciplines as well as in earth science data analysis, the intrinsic linear feature prevents this method from identifying nonlinear structures. This may be necessary as many geophysical phenomena are intrinsically nonlinear. As a consequence, many efforts have been made to extend PCA to grasp nonlinear relationships in data sets such as the principal curve theory (Hastie and Stuetzle, 1989) and the neutral network-based PCA (Kramer, 1991; Monahan, 2001), which is limited to low dimensional data or needs standard PCA for preprocessing. More recently, as the kernel method has been receiving growing attention in various communities, another nonlinear PCA implementation as a kernel eigenvalue problem has emerged (Schölkopf et al., 1998).

The kernel-based principal component analysis (KPCA) actually is implemented via a standard PCA in feature space, which is related to the original input space by a nonlinear implicit mapping (Schölkopf et al., 1998). KPCA has been recently applied to earth science data to explore nonlinear low dimensional structures (Tan, 2005; Tan et al., 2006). Ideally, the intrinsic nonlinear low dimensional structures in the data can be uncovered by using just a few nonlinear principal components. However, the dimension numbers of the feature space are usually much larger than the dimension numbers of the input space. Moreover, the variance cannot be quantitatively transferred from the feature space back into the input space. Consequently, the numbers of principal components which contribute meaningful amounts of variances are much larger than we commonly encounter in standard PCA results. Therefore, we need a mechanism to select nonlinear principal patterns and to construct the representative patterns. In this paper, we present the KPCA algorithms in language used in climate studies and propose a new KPCA pattern selection algorithm based on correlation with a natural phenomenon for KPCA applications to earth science data.

To the best of our knowledge, this work is the first and only effort on using KPCA in climate studies for knowledge acquisition. Therefore, in the following section, we first describe the PCA algorithm and then the KPCA algorithm in language comparable to the standard PCA applications in climate studies. Next, we present the newly proposed KPCA pattern selection algorithm. Then we briefly discuss the earth science data used for this work in Section 3 and describe the results in Section 4. In Section 5, we first discuss in-depth our understanding of the KPCA concepts, and finally present conclusions.

The contribution of this work includes two main points: 1) The emerging KPCA technique is described in the notation of PCA applications commonly used in earth science communities, and this work is the first KPCA application to earth science data; 2) A new spatial pattern selection algorithm based on correlation scores is developed here to overcome the problems of KPCA applications in earth science data sets, the overwhelming numbers of components and the lack of quantitative variance description.

2 ALGORITHM

KPCA emerged only recently from the well-known kernel theory, parallel to other kernel-based algorithms such as support vector machine (SVM) classification (Schölkopf et al., 1999a). In order to compare the similarities and the differences between standard PCA and KPCA, in this section, we first describe the commonly used standard PCA algorithms applied to earth science data analysis and then the KPCA algorithm for the same applications. Then, we discuss the limitation and new issues of KPCA such as pattern selection. Finally, we describe the new KPCA pattern selection algorithm.

2.1 PCA Algorithm

We follow the notations and procedures common to earth science communities to describe the standard PCA algorithm and a variant of its implementation by dual matrix (Emery and Thomson, 2001).

Suppose that we have a spatio-temporal data set, $\psi(\vec{x}_m, t)$, where $\vec{x}_m$ represents the given geolocation with $1 \leq m \leq M$, and t, the time, which is actually discretized at $t_n (1 \leq n \leq N)$. The purpose of the PCA as a data decomposition procedure is to separate the data into spatial parts $\phi(\vec{x}_m)$ and temporal parts $a(t)$ such that

$$\psi(\vec{x}_m, t) = \sum_{i=1}^{M} a_i(t)\phi_i(\vec{x}_m). \quad (1)$$

In other words, the original spatio-temporal data sets with N time snaps of spatial field values of dimension M can be represented by M number of spatial patterns. The contribution of those patterns to the original data is weighted by the corresponding temporal function $a(t)$. To uniquely determine the solution satisfying Equation (1), we place the spatial orthogonality condition on $\phi(\vec{x}_m)$ and the uncorrelated time variability condition on $a(t)$.

The above conditions result in an eigenvalue problem of covariance matrix C with λ being the eigenvalues and ϕ the corresponding eigenvectors. We can construct a data matrix D with

$$D = \begin{pmatrix} \psi_1(t_1) & \psi_1(t_2) & ... & \psi_1(t_N) \\ \psi_2(t_1) & \psi_2(t_2) & ... & \psi_2(t_N) \\ ... & ... & ... & ... \\ \psi_M(t_1) & \psi_M(t_2) & ... & \psi_M(t_N) \end{pmatrix}, \quad (2)$$

where $\psi_m(t_n) \equiv \psi(\vec{x}_m, t_n)$. Each row of matrix D is corresponding to one time series of the given physical values at a given location, and each column is a point in an M-dimensional space spanned by all locations, corresponding to a temporal snap. With the data matrix, the covariance matrix can be written as

$$C = fac * DD' \quad (3)$$

where the apostrophe symbol denotes the transpose operation. In the equation above, $fac = 1/N$.

Since the size of matrix D is $M \times N$, the size of matrix C is $M \times M$. Each eigenvector of C with M components represents a spatial pattern, that is, $\phi(\vec{x}_m)$. The corresponding time series (N values as a vector, $\vec{a}$) associated with a spatial pattern represented by $\overrightarrow{\phi_j}$ can be obtained by the projection of the data matrix onto the spatial pattern in the form of $\overrightarrow{a_j}' = \overrightarrow{\phi_j}' D$.

The advantage of the notations and procedures above, as normally defined in the earth science communities, is that they shed light on interpretation of the PCA results. Based on the matrix theory, the trace of covariance matrix C is equal to the sum of all eigenvalues. That is, $trace(C) \equiv \sum_{i=1}^{M} c_{ii} = \sum_{i=1}^{M} \lambda_i$. From Equation (2) and Equation (3), we have

$$c_{ii} = \frac{1}{N} \sum_{n=1}^{N} \left[\psi_i(t_n)\right]^2, \quad (4)$$

which is the the variance of the data at location $\vec{x}_i$ if we consider that the original data are centered against the temporal average (anomaly data). Therefore, the trace of C is the total variance of the data, and the anomaly data are decomposed into spatial patterns with corresponding time series for the contribution weights. The eigenvalues measure the contribution to the total variance by the corresponding principal components.

Computationally, solving an eigenvalue problem of an $M \times M$ matrix of DD' form is not always optimal. When $N < M$, the eigenvalue problem of matrix DD' can be easily converted into an eigenvalue problem of a dual matrix, $D'D$, of size $N \times N$ because the ranks and eigenvalues of DD' and $D'D$ are the same (Von Storch and Zwiers, 1999). Actually, the rank of the covariance matrix, r_C, is equal to or smaller than $min(M, N)$. The summation in Equation (1) and other corresponding equations should be from 1 to r_C instead of M.

The element of the dual matrix of the covariance matrix:

$$S = fac * D'D \quad (5)$$

is not simply the covariance between two time series. Instead,

$$s_{ij} = \frac{1}{N} \sum_{m=1}^{M} \left[\psi_m(t_i)\psi_m(t_j)\right], \quad (6)$$

can be considered as an inner product between two vectors which are denoted by columns in the data matrix D and are of M components. One should note that spatial averaging does make sense in earth science applications for one variable, unlike in traditional factor analysis. Nevertheless, the data values are centered against the temporal averages at each location. Due to this fact, the dual matrix S cannot be called a covariance matrix in strict meanings.

Since the matrix S is of size $N \times N$, the eigenvectors are not corresponding to the spatial patterns. Actually, they are corresponding to the temporal principal components, or the time series, $a(t)$. To obtain the corresponding spatial patterns, we need to project the original data (matrix D) onto the principal time series by

$$\overrightarrow{\phi_j} = D\overrightarrow{a_j} \quad (7)$$

with an eigenvalue dependent scaling.

2.2 KPCA Algorithms

In simple words, KPCA is the implementation of linear PCA in feature space (Schölkopf et al., 1998). With the same notation as we used in the previous section for the spatio-temporal data, we can recast the KPCA concept and algorithm as follows.

As in the case with dual matrix, we consider each snap of spatial field with M points as a vector of M components. Then, the original data can be considered as N M-dimensional vectors or N points in an M-dimensional space. Suppose that there is a map transforming a point from the input space (the space for original data) into a feature space, then we have

$$\Phi : R^M \to F; \vec{\psi} \mapsto \vec{X}. \quad (8)$$

Assume the dimension of the feature space is M_F, one vector in the input space, $\overrightarrow{\psi}_k$, is transformed into

$$\overrightarrow{X}_k \equiv \overrightarrow{\Phi(\overrightarrow{\psi}_k)} = \left(\Phi_1(\overrightarrow{\psi_k}), \Phi_2(\overrightarrow{\psi_k}), ..., \Phi_{M_F}(\overrightarrow{\psi_k})\right). \quad (9)$$

Similar to the data matrix in input space, we can denote the data matrix in the feature space as

$$D_\Phi = \begin{pmatrix} \Phi_1(\overrightarrow{\psi_1}) & \Phi_1(\overrightarrow{\psi_2}) & ... & \Phi_1(\overrightarrow{\psi_N}) \\ \Phi_2(\overrightarrow{\psi_1}) & \Phi_2(\overrightarrow{\psi_2}) & ... & \Phi_2(\overrightarrow{\psi_N}) \\ ... & ... & ... & ... \\ \Phi_{M_F}(\overrightarrow{\psi_1}) & \Phi_{M_F}(\overrightarrow{\psi_2}) & ... & \Phi_{M_F}(\overrightarrow{\psi_N}) \end{pmatrix}. \quad (10)$$

Unlike the standard PCA case, where we can actually solve an eigenvalue problem for either DD' or $D'D$ depending on the spatial dimension size and the number of observations (temporal size), we can only define

$$K = fac * D_\Phi{}' D_\Phi \tag{11}$$

for the eigenvalue problem in the feature space. This limitation comes from the so called "kernel trick" used for evaluating the elements of matrix K.

Comparing to the definition of s_{ij} for the standard PCA case, we will have the element of matrix K as

$$k_{ij} = fac*\left(D_\Phi{}' D_\Phi\right)_{ij} = \frac{1}{N}(\overrightarrow{\Phi(\vec{\psi}_i)} \bullet \overrightarrow{\Phi(\vec{\psi}_j)}). \tag{12}$$

The key of the kernel theory is that we do not need to explicitly compute the inner product. Instead, we define a kernel function for this product such that

$$k(\vec{x}, \vec{y}) = \left(\overrightarrow{\Phi(\vec{x})} \bullet \overrightarrow{\Phi(\vec{y})}\right). \tag{13}$$

Through the "kernel trick," we do not need to know either the mapping function Φ or the dimension size of the feature space, M_F, in all computations.

The main computation step in the KPCA is to solve the eigenvalue problem with $K\vec{\alpha} = \lambda\vec{\alpha}$. The eigenvalues still can be used to estimate the variance but only in the feature space. The eigenvector, as the case with dual matrix S in the standard PCA case, is playing a role of a time series. For the spatial patterns in the feature space, another projection, similar to that described in Equation (7),

$$\vec{v} = \sum_{m=1}^{M} \alpha_i \overrightarrow{\Phi(\vec{\psi}_i)}. \tag{14}$$

is needed. In practice, we do not need to compute $\vec{v}$ either. What we are more interested in is the spatial patterns we can obtain from the KPCA process. Therefore, we need to map back the structures represented by $\vec{v}$ in the feature space into the input space. Since the mapping from the input space to the feature space is nonlinear and implicit, it is not expected that the reverse mapping is simple or even unique. Fortunately, a preimage (data in the input space) reconstruction algorithm based on certain optimal condition has been developed already (Mika et al., 1999). In this process, all needed computations related to the mapping can be performed via the kernel function, and the algorithm is used in this work.

2.3 KPCA Pattern Selection Algorithm

Kernel functions are the key part in KPCA applications. There are many functions that can be used as kernel as long as certain conditions are satisfied (Schölkopf et al., 1998). Examples of kernel functions include polynomial kernels and Gaussian kernels (Schölkopf et al., 1999b). When the kernel function is nonlinear as we intend to choose, the dimension in the feature space is usually much higher than the dimension in the input space (Schölkopf et al., 1998). In special situations, the number of the dimensions could be infinite as in the case presented by the Gaussian kernel (Mika et al., 1999). The higher dimension in feature space is the desired feature for machine learning applications such as classification because data are more separated in the feature space and special characters are more easily identified. However, for spatial pattern extraction in earth science applications, the higher dimensionality introduces new challenges because we cannot simply pick one or a few spatial patterns associated with the largest eigenvalues.

Moreover, in standard PCA, the principal directions represented by the spatial patterns can be considered as the results of rotation of the original coordinate system. Therefore, the total variance of the centered data is conserved under the coordinate system rotation. As a result, significant spatial patterns are selected based on the contribution of variance by the corresponding patterns to the total variance. This can simply be calculated by the eigenvalues as discussed in Section 2.1. In the KPCA results, the mapping between the input space and the feature space is nonlinear. Therefore, the variance is not conserved from input space into the feature space. Consequently, although the eigenvalues still can be used to estimate the variance contribution in feature space, the variance distribution in the feature space cannot be quantitatively transferred back into variance distribution in the input space.

The introduction of higher dimensions in KPCA, that is, a large number of principal components and the difficulty to quantitatively describe the variance contribution in the input space by each component require a new mechanism for identifying the significant spatial patterns. A new pattern selection algorithm is developed (Tan, 2005) to overcome these problems as described below.

In standard PCA applications for earth science data analysis, the temporal components are usually correlated with a time series representing a given natural phenomenon. And the corresponding spatial pattern is claimed to be related to the phenomenon if the correlation coefficient is significantly different from zero. In KPCA, we cannot easily identify such spatial patterns, but we generally have more temporal components, as discussed in Section 2.2. The eigenvectors $\vec{\alpha}$, i.e., the KPCA temporal components, can be used to select KPCA components which can enhance the correlation to the given phenomenon.

After we perform the KPCA process on a particular set of data, we utilize an algorithm to obtain a

reduced set of temporal components in the pattern selection procedure. Although the variance in feature space does not represent the variance in the input space, we can still use the eigenvalues as a qualitative measurement to filter KPCA components which may contribute to the spatial patterns in the input space. We are interested in the significant KPCA components which are associated with, say, 99.9% variance in feature space as measured by the corresponding eigenvalues, and treat other components associated with very small eigenvalues as components coming from various noises. The algorithm sorts the temporal components in descending order according to their correlation score with a given phenomenon. Then linear combinations of them are tested from the highest score to the lowest, only the combinations that increase the correlation with the signal of interest are retained. The steps for combining the temporal components are:

- The correlation score of a vector **V** with the signal of interest is denoted as corr(**V**).
- Sort the normalized PC's according to the correlation score $\rightarrow PC_1, PC_2, \ldots, PC_p$.
- Save the current vector with the highest correlation score $\rightarrow \mathbf{V} := PC_1$.
- Save the current high correlation score as cc $\rightarrow cc := corr(PC_1)$.
- Maintain a list of the combination of sorted PC's $\rightarrow List\{\ \}$.

Where List.Add$\{1\}$ results in List$\{1\}$, List.Add$\{2\}$ results in List$\{1, 2\}$, etc...

Loop over the possible combinations of PC's that can increase the correlation score. If the score is increased, then keep the combination of PC's. The pseudo-code for the new pattern selection algorithm is given in Figure 1. In the pseuso-code and in the list above, p is the number of KPCA components we are interested in after de-noise.

```
V := PC_1
cc := corr(PC_1)
List.Add(1)
FOR i := 2 TO p
    IF corr(V + PC_i) > cc THEN
        V := V + PC_i
        cc := corr(V + PC_i)
        List.Add(i)
    END IF
END FOR
```

Figure 1: Pseudo-code of the pattern selection procedure.

The spatial patterns in input space are computed based on the preimage algorithm with all selected components (Mika et al., 1999).

3 DATA

A gridded global monthly Normalized Difference Vegetation Index (NDVI) data set was chosen to implement the KPCA. NDVI is possibly the most widely used data product from earth observing satellites. The NDVI value is essentially a measure of the vegetation greenness (Cracknell, 1997). As vegetation gains chlorophyll and becomes greener, the NDVI value increases. On the other hand, as vegetation loses chlorophyll, the value decreases.

The NDVI data used here were obtained from the NASA data web site (GES DISC (NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC)), 2006). The data are of $1^0 \times 1^0$ latitude-longitude spatial resolution with global coverage, and monthly temporal resolution with temporal coverage from January 1982 to December 2001. Since PCA analysis usually needs data without gaps, only data points with valid NDVI data in the whole period are chosen in the analysis. Therefore, we worked on global NDVI data for the 1982-1992 period only. Before using the data with PCA or KPCA, the NDVI data are deseasonalized by subtracting the climatological values from the original data. For that reason, the analysis is actually on NDVI anomalies.

In implementations, each point (location) in the physical coordinate system (the globe in latitude-longitude coordinates) is treated as one dimension, and time another dimension. Consequently, the data sets are represented in matrix format, and each column represents one month and each row element in the column represents a grid point value. In other words, all the latitude-by-longitude grid points for each month will be unrolled into one column of the data matrix. Therefore, the rows in each column represent a spatial location in latitude and longitude and each column represents a point in time as shown in the data matrix of Equation (2).

As a relationship between NDVI PCA patterns and El Niño Southern Oscillation (ENSO) was found (Li and Kafatos, 2000), we pick ENSO as the natural phenomenon for implementing the pattern selection algorithm. El Niño refers to a massive warming of the coastal waters of the eastern tropical Pacific. The Southern Oscillation refers to the fluctuations of atmospheric pressure in eastern and western Pacific (Philander, 1990), and its amplitude is described by a normalized sea level pressure difference between Tahiti and Darwin, also called Southern Oscillation Index (SOI) (Ropelewski and Jones, 1987). Because El Niño is highly correlated with one phase of the southern oscillation, the phenomenon is usually called El Niño Southern Oscillation (ENSO). ENSO is the largest known global climate variability on interannual timescales, and the SOI is one of the representative signals of ENSO. The SOI represents

a climatic anomaly that has significant global socio-economic impacts including flooding and drought pattern modification. The SOI data used here were obtained from NOAA National Weather Service, Climate Prediction Center (CPC (Climate Predication Center/NOAA), 2006).

4 RESULTS

The standard linear PCA is first used to the spatio-temporal NDVI anomaly data. As a widely used procedure, we correlate the principal temporal components with the SOI time series and find that the correlation is strongest between the fourth component (the component corresponding to the fourth largest eigenvalue) and SOI. The correlation coefficient is 0.43, and this component contributes 3.8% of the total variance. The corresponding simple spatial pattern is displayed in Figure 2.

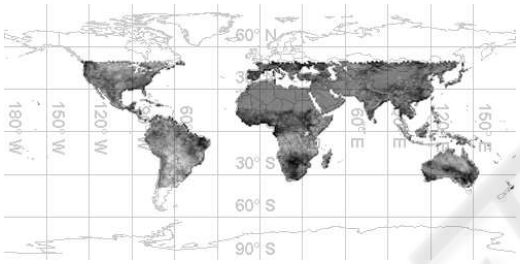

Figure 2: Simple NDVI spatial pattern of the fourth spatial component from standard PCA. The gray scale denotes the anomaly values. The darkest is corresponding to the highly positive anomaly values.

In the KPCA analysis, with trials of several kernels for the best results, we choose the Gaussian kernel,

$$k(\vec{x},\vec{y}) = exp\left(-\frac{\parallel \vec{x}-\vec{y} \parallel^2}{2\sigma^2}\right). \qquad (15)$$

for the demonstration. We then use the pattern selection algorithms described in Section 2.3 to obtain a combined spatial pattern. In order to attain a high correlation, the free parameter σ in the Gaussian kernel had to be adjusted. Using the data set's standard deviation for σ in the Gaussian kernel did not produce the best results. It is possible that the kernel under-fits the data with that σ. A σ being equal to 26% of the standard deviation of the NDVI data set resulted in the correlation score with SOI of $r = 0.68$. Twenty (20) of the 131 eigenvectors were used, and those are about 15% of the significant KPCA components. The corresponding combined spatial pattern with those selected components is presented in Figure 3.

For comparison, the same pattern selection algorithm is also applied to the standard PCA results. In this case, 28 of 120 eigenvectors are selected for enhancing the correlation set initially by the fourth component. The resulting correlation coefficient is $r = 0.56$. The corresponding combined spatial pattern based on the 28 selected PCA components is demonstrated in Figure 4. Apparently, the pattern selection algorithm is more efficient and effective with the KPCA application than with the standard PCA application because we achieve higher correlation scores with fewer components in the KPCA case than in the standard PCA case.

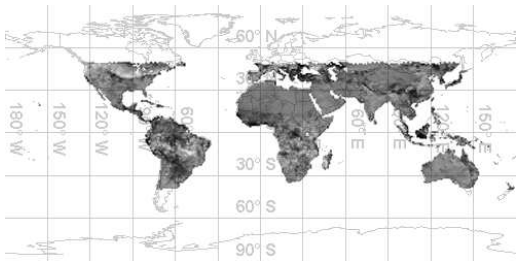

Figure 3: Combined NDVI spatial pattern from KPCA results based on Gaussian kernel. The gray scale is the same as that in Figure 2.

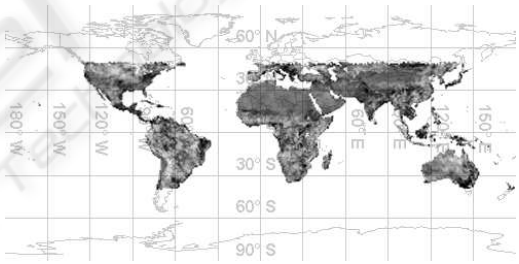

Figure 4: Combined NDVI spatial pattern from standard PCA results with the same pattern selection algorithm as for KPCA. The gray scale is the same as that in Figure 2.

By comparing Figure 4 and Figure 3 against Figure 2, we can find that the combined patterns from either the standard PCA or KPCA components show higher-resolution structure than the simple pattern presented by a single PCA component. This result is not unexpected because PCA components contain high resolution information in components with low eigenvalues. In other words, the first principal component associated with the largest eigenvalue catches large scale features of the data. The key point is that with standard principal component analysis, we can only pick one component to be associated with a given phenomenon through a correlation analysis. Once the component is identified, we cannot associate other components to the same phenomenon. The pattern selection algorithm described in this paper provides a mechanism to select multiple principal components

with one phenomenon.

To explore the difference for information extraction from the combined patterns and the simple pattern, we display a world drought map for the 1982-1983 El Niño episode in Figure 5 (NDMC (National Drought Mitigation Center), 2006) because the NDVI dataset used here spans the 1982-1992 period. Please note that the correlation selection in our case is based on a positive correlation coefficient while the values of SOI associated with El Niño are negative. Therefore, in the spatial patterns based on an NDVI anomaly (Figres 2–4), positive values are actually associated with a negative NDVI anomaly due to ENSO, which in return, is associated with the drought patterns in Figure 5.

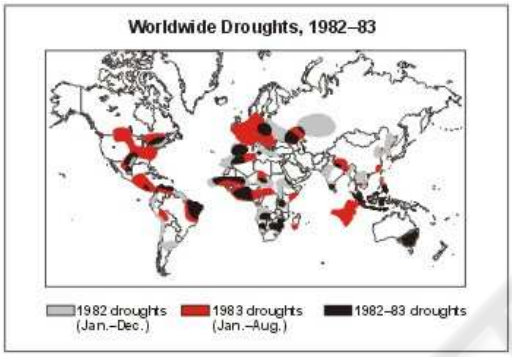

Figure 5: World drought pattern during the 1982-1983 El Niño episode (from the web site of National Drought Mitigation Center (NDMC (National Drought Mitigation Center), 2006)).

The simple PCA pattern (Figure 2) does capture drought patterns, but in large scale only, such as droughts in the Amazon area, southern Africa, and Australia in the 1982-1983 period. However, the shapes and sizes of the drought patterns are difficult to compare with the simple PCA pattern. In contrast, the combined patterns from the selection algorithm applications on standard PCA and KPCA capture the details such as the curvature in the drought patterns in the continental US for the 1983 drought. The combined KPCA pattern also shows good agreement on the drought patterns in western Africa around Ivory Coast. The drought pattern in Malaysia and Borneo Island (around 112E longitude near the Equator) in the South & East Asia region is evident in the combined patterns from both standard PCA and KPCA, but they are not exhibited in the simple PCA pattern. Another apparent improvement from the combined KPCA spatial pattern is that the drought in Europe is more accurately identified in contrast to the simple PCA pattern.

5 DISCUSSION AND CONCLUSIONS

From a data decomposition perspective, PCA as well as KPCA are data adaptive methods. That means that the bases for the decomposition are not chosen *a priori*, but are constructed from the data. In the standard linear PCA case, the orthogonality condition on the spatial patterns and the uncorrelated condition on temporal components guarantee the uniqueness of the decomposition. Additional freedoms introduced by the implicit nonlinear transformation make the uniqueness condition invalid, and the KPCA results depend on the nonlinear structure implicitly described by the kernel. As a result, different kernels should be tested before significant results can be discovered because the underlying nonlinear structure can only be picked up by a kernel with a similar structure.

In a broad sense, principal component analysis describes the co-variability among multivariate observations. The definition of the co-variability between two observations actually determines the core structure one may expect from the result. The most commonly used definition is covariance or correlation between points defined in either object space or variable space (Krzanowski, 1988). In KPCA application, if we do not consider the process as a mapping from input space into the feature space, we can treat the "kernel trick" as another definition of the pair-wise co-variability. However, this definition of the co-variability can only be implemented on data points defined for each observation. That is, the KPCA is applied to object space only. This results in the eigenvalue problem for KPCA being always on a matrix of size $N \times N$, even when M, the number of variables or geolocations in earth science applications is smaller than N. Since the mapping function Φ is never determined in the procedure, the computationally efficient SVD procedure cannot be used either, because the data matrix in feature space, D_Φ, is not known.

The pair-wise co-variability is actually a measure of the pair-wise proximities. Therefore, KPCA can be understood in a broad sense as a general means to discover "distance" or "similarity (dissimilarity)" based structure. That is why most dimension reduction algorithms such as Multidimensional Scaling (MDS) (Cox and Cox, 2000), Locally Linear Embedding (LLE) (Roweis and Saul, 2000), and Isomap (Tenenbaum et al., 2000) can be related to KPCA algorithm (Ham et al., 2004).

In conclusion, the KPCA algorithm is recast in the notation of PCA commonly used in earth science communities and is used for NDVI data. To overcome the problems of KPCA applications in earth sciences, namely the overwhelming numbers of components and lack of quantitative variance description,

a new spatial pattern selection algorithm based on correlation scores is proposed here. This selection mechanism works both on standard PCA and KPCA, and both give superior results compared to the traditional simple PCA pattern. In the implementation example with NDVI data and the comparison with the global drought patterns during the 1982-1983 El Niño episode, the combined patterns show much better agreement with the drought patterns on details such as locations and shapes.

REFERENCES

Cox, T. F. and Cox, M. A. (2000). *Multidimensional Scaling*. Chapman & Hall.

CPC (Climate Predication Center/NOAA) (2006). (STAND TAHITI - STAND DARWIN) SEA LEVEL PRESS ANOMALY. http://www.cpc.ncep.noaa.gov/data/indices/soi (Last accessed on Feb. 5, 2006).

Cracknell, A. P. (1997). *The Advanced Very High Resolution Radiometer*. Taylor & Francis Inc.

Emery, W. J. and Thomson, R. E. (2001). *Data Analysis Methods in Physical Oceanography*. Elsevier.

GES DISC (NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC)) (2006). Pathfinder AVHRR Land Data. ftp://disc1.gsfc.nasa.gov/data/avhrr/Readme.pal (Last accessed on Feb. 9, 2006).

Haddad, R. A. and Parsons, T. W. (1991). *Digital Signal Processing: Theory, Applications, and Hardware*. Computer Science Press.

Ham, J., Lee, D., Mika, S., and Schölkopf, B. (2004). Kernel View of the Dimensionality Reduction of Manifolds. In *Proceedings of the 21st International Conference on Machine Learning*.

Hastie, T. and Stuetzle, W. (1989). Principal Curves. *Journal of the American Statistical Association*, 84:502–516.

Holmes, P., Lumley, J. L., and Berkooz, G. (1996). *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press.

Kramer, M. (1991). Nonlinear Principal Component Analysis Using Autoassociative Neural Networks. *AIChE J.*, 37(2):233–243.

Krzanowski, W. J. (1988). *Principles of Multivariate Analysis: A User's Perspective*. Oxford University Press.

Li, Z. and Kafatos, M. (2000). Interannual Variability of Vegetation in the United States and Its Relation to El Niño/Southern Oscillation. *Remote Sensing of Environment*, 71(3):239–247.

Lorenz, E. N. (1959). Empirical orthogonal functions and statistical weather prediction. Final Report, Statistical Forecasting Project, 1959; Massachusetts Institute of Technology, Dept. of Meteorology, 29–78.

Mika, S., Schölkopf, B., Smola, A., Müller, K.-R., Scholz, M., and Rätsch, G. (1999). Kernel PCA and Denoising in Feature Spaces. In Kearns, M. S., Solla, S. A., and Cohn, D. A., editors, *Advances in Neural Information Processing Systems 11*, pages 536 – 542, Cambridge, MA. MIT Press.

Monahan, A. H. (2001). Nonlinear Principal Component Analysis: Tropical IndoPacific Sea Surface Temperature and Sea Level Pressure. *Journal of Climate*, 14(2):219–233.

NDMC (National Drought Mitigation Center) (2006). What is Drought? http://www.drought.unl.edu/whatis/elnino.htm (Last accessed on Feb. 8, 2006).

Philander, S. G. (1990). *El Niño, La Niña, and the Southern Oscillation*. Academic Press.

Ropelewski, C. and Jones, P. (1987). An Extension of the Tahiti–Darwin Southern Oscillation Index. *Monthly Weather Review*, 115(9):2161–2165.

Roweis, S. and Saul, L. (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(22 December 2000):2323–2326.

Schölkopf, B., Burges, C., and Smola, J. (1999a). *Advances in Kernel Methods: Support Vector Learning*. MIT Press.

Schölkopf, B., Mika, S., Burges, C., Knirsch, P., Müller, K.-R., Rätsch, G., and Smola, A. (1999b). Input Space vs. Feature Space in Kernel-Based Methods. *IEEE Transactions on Neural Networks*, 10(5):1000–1017.

Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10(5):1299–1319.

Tan, J. (2005). *Applications of Kernel PCA Methods to Geophysical Data*. George Mason University. PhD Thesis.

Tan, J., Yang, R., and Kafatos, M. (2006). Kernel PCA Analysis for Remote Sensing Data. In *18th Conference on Climate Variability and Change*. American Meteorological Society. Paper P1.5, Altanta, GA, CD-ROM.

Tenenbaum, J., de Silva, V., and Langford, J. (2000). A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(22 December 2000):2319–2323.

Thompson, D. W. J. and Wallace, J. M. (2000). Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability. *Journal of Climate*, 13(5):1000–1016.

Von Storch, H. and Zwiers, F. W. (1999). *Statistical Analysis in Climate Research*. Cambridge University Press.

Wallace, J. M., Smith, C., and Bretherton, C. S. (1992). Singular Value Decomposition of Wintertime Sea Surface Temperature and 500-mb Height Anomalies. *Journal of Climate*, 5(6):561–576.